March 2, 2014

Chairman Tom Wheeler
Commissioner Mignon Clyburn
Commissioner Jessica Rosenworcel
Commissioner Ajit Pai
Commissioner Michael O'Rielly
Federal Communications Commission
445 Twelfth Street SW
Washington, DC 20054

Vitaly Shmatikov
Associate Professor
Department of Computer Science
The University of Texas at Austin
Austin, TX 78712

Re: WC Docket No. 13-306, *Public Knowledge Petition for Declaratory Ruling*

Dear Chairman Wheeler and Commissioners:

I am an Associate Professor in the Department of Computer Science at the University of Texas at Austin, one of the top ten highest ranked computer science departments in the U.S. My area of expertise is computer security and privacy. I received the 2008 PET Award for Outstanding Research in Privacy-Enhancing Technologies for my work on data anonymization and re-identification. I was also the runner-up for the 2013 PET Award. I am a recipient of the CAREER Award from the National Science Foundation (a prestigious grant that supports my research on digital privacy). My research papers on anonymity and privacy received multiple awards, including the Best Practical Paper Award from the IEEE Symposium on Security and Privacy and the NYU-Poly AT&T Best Applied Security Paper Award.

Anonymized CDRs (Call Detail Records) Can Be Re-Identified

Anonymization may remove direct identifiers from phone-call records, but the remaining information can still be used learn enough about specific individuals to re-identify them. De-identified records retain enough information to (1) link all calls made by the same individual, (2) deduce the approximate location of the individual at the time of each call (based on the nearest cell tower, whose location is included in the anonymized CDRs), and (3) learn other data related to the individual, such as the time and duration of each call.

Multiple scientific studies have reached the conclusion that **anonymization of CDRs does not work**. For example, a recent large-scale study by Zang and Bolot of more than 30 billion call records made by 25 million cell-phone users across all 50 states of the U.S. demonstrated that as many as 60% of these users can be re-identified by linking the location information in their CDRs to U.S. Census records and other publicly available sources of information.[1]

The process of re-identification typically starts by using the anonymized CDRs to deduce the geographic locations of anonymous callers at certain points in time. Isaacman et al. demonstrated that anonymized CDRs can be used with high accuracy to determine important places in people's lives.[2] These places include individuals' homes, workplaces, houses of worship, and other locations where they spend a significant amount of time and/or visit frequently.

Simply knowing a person's home and work locations at the granularity of a census block (which in many urban areas is at least as big as the area covered by a single cell tower and thus can be learned from the anonymized CDRs) is sufficient to uniquely re-identify a median person in the U.S. working population.[3] Even knowledge of the approximate locations at the granularity of a census tract is sufficient to reduce the "anonymity set" (i.e., the number of individuals in the U.S. working population to whom a particular anonymized CDR might belong) to 21 individuals. People whose homes and workplaces are located in two different regions can be re-identified even more easily.

Anonymized CDRs reveal much more than home/work location pairs. By linking calls made by the same individual and the locations of the corresponding cell towers obtained from the CDRs, it is very easy to reconstruct the entire "trajectory" taken by this individual throughout the day. These trajectories, also known as "mobility traces" or "mobility patterns," include, for example, the route of the person's highway commute and the path taken when walking his or her children to school. Mobility traces are even easier to re-identify than simple home/work location pairs, especially with the additional information such as the time of each call and the knowledge of other important locations in the person's life, such as their gym, favorite restaurant, or place of worship, which – as demonstrated by Isaacman et al. – can also be deduced from the anonymized CDRs. With just a little bit of additional information, an average

---

[1] Hui Zang, Jean Bolot: Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study. In Proceedings of the 17th Annual International Conference on Mobile Computing and Networking (MobiCom 2011), pp. 145-156, 2011.

[2] Sibren Isaacman, Richard A. Becker, Ramón Cáceres, Stephen G. Kobourov, Margaret Martonosi, James Rowland, Alexander Varshavsky: Identifying Important Places in People's Lives from Cellular Network Data. In Proceedings of the 9th International Conference on Pervasive Computing (Pervasive 2011), pp. 133-151, 2011.

[3] Philippe Golle, Kurt Partridge On the Anonymity of Home/Work Location Pairs. In Proceedings of the 7th International Conference on Pervasive Computing (Pervasive 2009), pp. 390-397, 2009.

mobility trace (i.e., a sequence of CDRs belonging to the same individual but taken at different locations) can be re-identified with high confidence.[4] [5]

In summary, **anonymized CDRs containing location information are highly vulnerable to re-identification** and are likely to be re-identified if released.

Even Without Re-Identification, Anonymized CDRs Reveal Sensitive Information about Individuals

Anonymized CDRs are a rich source of information about individuals. By analyzing the network structure of the calls (i.e., who called whom, how frequently, etc.) and geographic "coincidences" (i.e., two or more individuals making calls from the same location at approximately the same time), it is possible to deduce social ties between individuals[6] and to determine which individuals belong to the same group (for example, enrolled in the same middle school).[7]

Social ties and group membership are highly sensitive from the privacy perspective. They may reveal religious affiliation, political ties, sexual orientation, and other private information. If a member of such a group is ever re-identified, not only will his identity be revealed, but also his affiliations and preferences. Furthermore, even without identifying individual group members, anonymized CDRs can be used to accurately estimate the number of people attending a particular church, political meeting, etc.

Finally, simply knowing that an anonymous person belongs to certain groups can be sufficient to re-identify this person if group directories are public – for example, if the groups have a presence on an online social networking site such as Facebook. If the adversary knows that some anonymous individual is a member of groups A, B, and C, computing the intersection of these groups' memberships can pinpoint this individual with high accuracy.[8]

---

[4] Chris Y.T. Ma, David K.Y. Yau, Nung Kwan Yip, Nageswara S.V. Rao: Privacy Vulnerability of Published Anonymous Mobility Traces. In Proceedings of the 16th Annual International Conference on Mobile Computing and Networking (MobiCom 2010), pp. 185-196, 2010.

[5] Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen, Vincent D. Blondel: Unique in the Crowd: The Privacy Bounds of Human Mobility. In Nature Scientific Reports 3, 1376; 2013.

[6] David J. Crandall, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, Jon Kleinberg: Inferring Social Ties from Geographic Coincidences. In Proceedings of the National Academy of Sciences, 107(52), 2010.

[7] Richard A. Becker, Ramón Cáceres, Karrie Hanson, Ji Meng Loh, Simon Urbanek, Alexander Varshavsky, Chris Volinsky: Clustering Anonymized Mobile Call Detail Records to Find Usage Groups. In Proceedings of the 1st Workshop on Pervasive Urban Applications, 2011.

[8] Gilbert Wondracek. Thorsten Holz, Engin Kirda, Christopher Kruegel: A Practical Attack to De-Anonymize Social Network Users. In Proceedings of the 31st IEEE Symposium on Security and Privacy, pp. 223-238, 2010.

In general, anonymizing a dataset by simply removing "personally identifiable information" is not sufficient to guarantee privacy of the individuals whose records appear in the dataset. As long as any of the data elements remaining in the records can be used to distinguish between anonymized individuals – for example, locations and times of their phone calls, relationships between calls, etc. – these elements can help link anonymized records to public sources of identifiable information and to re-identify at least some of the anonymized records.[9]

Conclusion

Anonymized call detail records (CDRs) are highly vulnerable to re-identification, especially if they contain information about the approximate location and time of each call. Multiple scientific studies, including a large-scale study by Zang and Bolot of more than 30 billion call records, demonstrated that an average user whose records appear in such a dataset can be re-identified with high probability. Even without re-identification, anonymized CDRs reveal sensitive information such as the size of membership in political, religious, and social groups.

In my opinion, releasing anonymized CDRs presents a significant privacy risk to the individuals whose data appears in these CDRs.

Respectfully submitted,

Vitaly Shmatikov

Associate Professor
Department of Computer Science
The University of Texas at Austin
Austin, TX 78712

[9] Arvind Narayanan, Vitaly Shmatikov. "Myths and Fallacies of 'Personally Identifiable Information'". Communications of the ACM, 53(6), pp. 24-26, 2010.